ARTICLE

Seasonal influence on respiratory tract infection severity including COVID-19 quantified through Markov Chain modeling

Rob C. van Wijk[1] 🆔 | **Laurynas Mockeliunas**[1] 🆔 | **Caryn M. Upton**[2] |
Jonathan Peter[3] | **Andreas H. Diacon**[2] | **Ulrika S. H. Simonsson**[1] 🆔

[1]Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden

[2]TASK, Cape Town, South Africa

[3]Department of Medicine, University of Cape Town Lung Institute and Division of Allergy and Clinical Immunology, University of Cape Town, Cape Town, South Africa

Correspondence
Ulrika S. H. Simonsson, Box 591, Uppsala 75124, Sweden.
Email: ulrika.simonsson@farmbio.uu.se

Abstract

Respiratory tract infections (RTIs) are a burden to global health, but their characterization is complicated by the influence of seasonality on incidence and severity. The Re-BCG-CoV-19 trial (NCT04379336) assessed BCG (re)vaccination for protection from coronavirus disease 2019 (COVID-19) and recorded 958 RTIs in 574 individuals followed over 1 year. We characterized the probability of RTI occurrence and severity using a Markov model with health scores (HSs) for four states of symptom severity. Covariate analysis on the transition probability between HSs explored the influence of demographics, medical history, severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2), or influenza vaccinations, which became available during the trial, SARS-CoV-2 serology, and epidemiology-informed seasonal influence of infection pressure represented as regional COVID-19 pandemic waves, as well as BCG (re)vaccination. The infection pressure reflecting the pandemic waves increased the risk of RTI symptom development, whereas the presence of SARS-CoV-2 antibodies protected against RTI symptom development and increased the probability of symptom relief. Higher probability of symptom relief was also found in participants with African ethnicity and with male biological gender. SARS-CoV-2 or influenza vaccination reduced the probability of transitioning from mild to healthy symptoms. Model diagnostics over calendar-time indicated that COVID-19 cases were under-reported during the first wave by an estimated 2.76-fold. This trial was performed during the initial phase of the COVID-19 pandemic in South Africa and the results reflect that situation. Using this unique clinical dataset of prospectively studied RTIs over the course of 1 year, our Markov Chain model was able to capture risk factors for RTI development and severity, including epidemiology-informed infection pressure.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

Respiratory tract infections (RTIs) are subject to seasonal influence on their occurrence and severity.

WHAT QUESTION DID THIS STUDY ADDRESS?

Can we develop an epidemiology-informed pharmacometric model to quantify RTI occurrence and severity, including the effect of infection pressure informed by coronavirus disease 2019 (COVID-19) cases?

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

We have integrated epidemiological data into the pharmacometric modeling workflow to quantify seasonal influence on RTIs. Higher COVID-19 case numbers increased incidence and severity of RTIs, and decreased RTI transition to milder symptoms. African ethnicity, male biological gender, and severe acute respiratory syndrome-coronavirus 2 seropositivity protect against RTI worsening. Our integrated method enabled estimation of the under-reporting of COVID-19 cases during the first wave.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

Including both time-in-trial and calendar-time in model development and diagnostics is recommended when the topic of study is subject to seasonal influence.

INTRODUCTION

Respiratory tract infections (RTIs) are a burden to global health. In the United States alone, an estimated 35 million symptomatic cases of influenza occurred in 2019–2020, with a mortality of 20,000.[1] Globally, up to 4 million patients died annually due to respiratory diseases.[2] The challenge in studying RTIs is their seasonality. Annual epidemics, such as influenza, have seasonal patterns influencing incidence based on the geographic location.[3–5] The recent coronavirus disease 2019 (COVID-19) pandemic caused by the severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2) pathogen, like the earlier SARS and Middle East respiratory syndrome (MERS) pandemics, followed a similar pattern, albeit less dependent on season and more on individual and population behavior.[6–9] A quantitative model-based approach, such as a pharmacometric model informed by epidemiological data,[10] is therefore warranted to understand RTI patterns over time, and to distinguish vaccine or treatment effects on the course of the infection.

The Re-BCG-CoV-19 trial (NCT04379336) evaluated the influence of Bacillus Calmette-Guérin (BCG) (re)vaccination of healthcare workers on COVID-19 morbidity and mortality in South Africa.[11] This phase III clinical trial followed 1000 healthcare workers for 1 year and recorded occurrence and severity of RTIs with a weekly resolution, providing a uniquely large dataset to study RTIs across different seasons, pandemic waves, and vaccine interventions, within a diverse population. Incidence of RTIs was a secondary study end point, and severity was reported on a 7-point scale ranging from mild, moderate, and severe symptoms, to hospitalization, hospitalization with supplemental oxygen, hospitalization with mechanical ventilation, and death.

Longitudinal analysis of clinical trials is commonly performed over time-in-trial or time-since-intervention. However, to study seasonal influence on the trial's end point, the calendar-time needs to be taken into consideration as well. The Re-BCG-CoV-19 trial data, which was collected over a year for each participant, could be leveraged for this purpose. Half of the trial participants had been enrolled by July 1, 2020, in the middle of the typical Southern Hemisphere influenza season,[12] with 95% of enrollment spanning half a year (May 6, 2020 to October 12, 2020). Seasonal influence on the incidence and severity of RTIs could therefore be quantified by utilizing the epidemiological data on infection pressure corresponding to the calendar time as a time-varying covariate.

The objective of this work was to characterize the probability of RTI occurrence and severity, and probability of change in severity, using a Markov model, which was used to investigate the impact of covariates on transition probability, including epidemiology-informed infection pressure in the regions where the trial was conducted. Finally, the effect of BCG (re)vaccination on the transition probabilities was assessed.

METHODS

Data

The Re-BCG-CoV-19 trial (NCT04379336) was a double-blind, randomized, placebo-controlled phase III trial conducted at three sites in the Western Cape, South Africa. Two sites were in the City of Cape Town and one site in George, a district capital approximately 400 km to the East of Cape Town. One thousand healthcare workers were enrolled between May 4 and October 23, 2020, and vaccinated 1:1 with either BCG or placebo. At baseline, week 10, week 26, and week 52, participants were tested for SARS-CoV-2 antibodies through immunoglobulin G serology. Tuberculosis (TB) infection status was tested by interferon-gamma release assay (IGRA) at baseline and week 52. SARS-CoV-2 specific vaccines became available to the study population from February 2021. Therefore, the time of administration of a vaccine other than BCG was recorded and included in the analysis as either a covariate on the transition probabilities between health score (HS) states, or a reason for censoring of the data. The analysis was performed on the intention-to-treat dataset, in which participants were censored at their final visit at 52 weeks, or at loss to follow-up, or death. Informative dropout was considered in the analysis unless less than 5% of participants were censored prior to their final visit. Further details on the trial methods, primary outcome, and protocol are available in the original publication.[11]

RTI events were recorded on a 7-point HS (1: mild, 2: moderate, 3: severe, 4: hospitalization, 5: hospitalization with supplemental oxygen, 6: hospitalization with mechanical ventilation, and 7: death), this HS represented the state of the participant. If no RTI event was recorded for an individual, the HS 0 (healthy) was assumed for that time. Grouping of HSs was considered in case of limited data on one or more HSs. RTIs were categorized following the Medical Dictionary for Regulatory Activities (MedDRA) terminology and documented with a weekly resolution. Two time variables were recorded: the time-in-trial and the calendar-time to correlate events to the pandemic waves.

Epidemiology-informed infection pressure was accounted for by utilizing reported COVID-19 cases per capita in the Western Cape, South Africa, to account for the seasonal wave pattern of the COVID-19 epidemic and its impact on RTI events in the population. Regional (Cape Town and George) COVID-19 cases per capita data were matched to the study site to improve geographic resolution, as wave onset and magnitude were different between regions within the Western Cape province.[13]

Model development

A Markov Chain model was developed, quantifying the probabilities of transitioning between states which represented the HSs for each RTI at each timepoint (Figure 1a). Model development was performed first structurally to define the number of states and transitions. Subsequently, an interindividual variability (IIV) model was developed, followed by covariate modeling, to account for predictive covariates on the transitions, including epidemiology-informed infection pressure. In the case of multiple events per timepoint per individual, inclusion of inter-event variability was considered. Lastly, the effect of BCG (re)vaccination as the study intervention was tested.
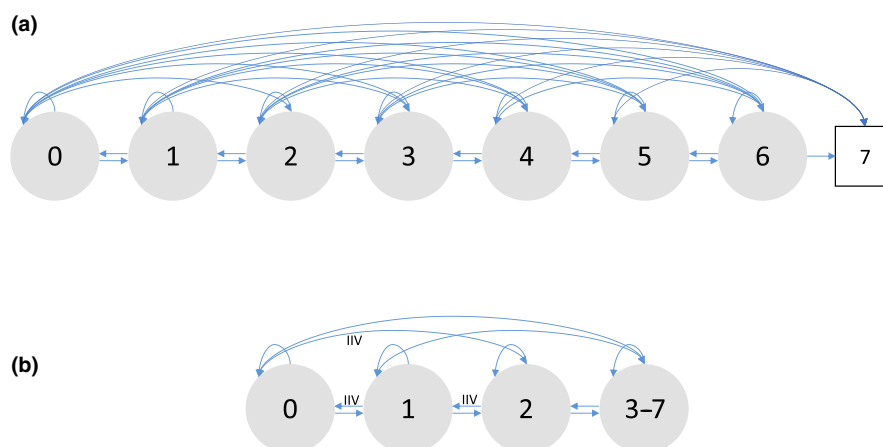

FIGURE 1 (a) Theoretical Markov Chain model with 8 health score states representing each of the health scores 0–7. Arrows depict transitions between health scores which are depicted as circles except for health score 7 (death) from which no transitions are possible. (b) Final structural and interindividual variability (IIV) model with grouped health scores 3–7 into a single state, and IIV terms on the transition probabilities between health score states 1 and 0, 2 and 1, and 2 and 0, respectively.

Probability parameters were logit transformed (Equation 1) to allow for unconstrained parameter estimation and inclusion of IIV terms and covariate relationships. Logistic transformation (Equation 2) was performed to report probabilities between [0,1] for interpretability.[14]

$$PL_{i \to j} = \log\left(\frac{\theta_{i \to j}}{1 - \theta_{i \to j}}\right) \tag{1}$$

$$P_{i \to j} = \frac{e^{PL_{i \to j}}}{1 + e^{PL_{i \to j}}} \tag{2}$$

where $PL_{i \to j}$ is the logit transformed probability of transitioning from state i to state j conditional on not transitioning to another state, $\theta_{i \to j}$ is the probability parameter estimated by the model, and $P_{i \to j}$ is the probability of transitioning from state i to state j conditional on not transitioning to another state constrained to [0,1]. The sum of the probabilities of all transitions from a single state equals 1. Thus, for each state, one transition probability was estimated, for example, $P_{1 \to 0}$, whereas the other transition probabilities were calculated from estimates of the probabilities conditional on not transitioning to another state, for example, $T_{1 \to 2} = P_{1 \to 2} * (1 - P_{1 \to 0})$, to ensure the sum would always equal 1.

The IIV model was developed by univariably testing IIV terms on each transition probability on the logit domain (Equation 3).

$$PL_{i \to j,k} = PL_{i \to j} + \eta_k + \text{cov}_l + E_{BCG} \tag{3}$$

where $PL_{i \to j,k}$ is the logit transformed transition probability from state i to state j for individual k conditional on not transitioning to another state, $PL_{i \to j}$ is the logit transformed population transition probability from state i to state j conditional on not transitioning to another state, η_k is the individual deviation from the population parameter value for individual k drawn from a normal distribution with mean 0 and variance ω^2, cov_l is the covariate relationship for covariate l, and E_{BCG} is the effect of BCG on the transition probability. Statistically significant ($p < 0.05$) IIV terms were included in the model in a stepwise manner until no further statistical improvement of the model resulted. After the final IIV model was developed, the covariate analysis was performed, following the stepwise covariate modeling (SCM) procedure, testing linear, piecewise linear with two slopes, power, and exponential functions.[15] The following covariates were considered: age, body mass index (BMI), biological gender, self-reported ethnicity, job category, SARS-CoV-2 serology at baseline, IGRA-based TB status at baseline, TB IGRA conversion during trial, medical history of hypertension, medical history of asthma, medical history of chronic obstructive pulmonary disease (COPD), medical history of other lung diseases, smoking (both categorical yes/no, and continuous pack years), expected interaction with patients with COVID-19, and the time varying covariates of reported COVID-19 cases to inform on the epidemiology-informed infection pressure, SARS-CoV-2 specific and influenza vaccination, and SARS-CoV-2 antibody seroconversion during the trial. Potential covariate relationships for demographic, social, or medical risk factors were included univariably (Equation 3). Epidemiology-informed infection pressure, reflecting the pandemic waves, was tested with a linear, power, or exponential function. Potential covariates were considered for inclusion if they met the cutoff for statistical significance ($p < 0.05$ for forward inclusion). At each step, the most statistically significant covariate relation was added to the covariate model until no more statistical improvement of the model was observed, after which backward deletion steps followed ($p < 0.01$ for backward deletion) until the final covariate set was obtained. The final covariate model was evaluated on both time-in-trial and calendar-time (see Model Evaluation below), with the purpose to correct any misfits on either timescale and improve the model through the inclusion of a time-specific covariate. Lastly, the impact of BCG versus the control arm was tested on the covariate model which controlled for all other risk factors to test the effect of the intervention without any confounding covariate effects (Equation 3).

Model evaluation

Model performance was evaluated numerically and graphically. Numerically, pharmacologically interpretable parameter estimates as well as precision of parameter estimates reported in relative standard error (RSE) were considered. The objective function value (OFV) was utilized to test for statistical significance levels using the likelihood ratio test for nested models assuming a χ^2-distribution, where for 1 degree of freedom, a drop in OFV of 3.84 corresponded to $p < 0.05$ and a drop in OFV of 6.63 corresponded to $p < 0.01$. For models that were not nested, the Akaike Information Criterion (AIC) was utilized to inform on model selection.[16]

Graphically, diagnostics were simulation-based as Markov Chain model predicts probabilities of an individual being in a given state which prevents the use of goodness of fit plots conventionally used in pharmacometrics.[17] Two graphical diagnostics over time were developed, both over time-in-trial and calendar-time, to evaluate the model performance related to the time since vaccination as well as to the seasonal influence of infection pressure on RTIs. Both diagnostics are visualized with a weekly resolution. For each simulation, a state was assigned to an individual at each timepoint, using the estimated probabilities and a random number generator. Summary statistics on the proportion of the simulated population in each state relative

to the total population in the trial at the corresponding timepoint were calculated. This was graphically overlaid with the proportion in the corresponding state as observed in the trial, to assess the model performance.

Nonlinear mixed effects modeling was performed using the Laplacian method in NONMEM[18] (version 7.5.0) assisted by PsN[19,20] (version 5.2.6). A reproducible workflow was developed for data handling and graphics throughout the trial,[21] which was performed in R[20] (version 4.0.4) through the RStudio[22] (version 1.4.1106) user interface.

RESULTS

Trial overview

Throughout the trial, 958 RTI events were recorded in 574 individuals with an average duration of 2.2 weeks. The remaining individuals ($n = 426$) were retained in the analysis dataset with a HS of 0 reflecting their absence of RTIs. The RTIs were most commonly reported as respiratory tract infection (39.0%), COVID-19 (22.4%), upper respiratory tract infection (15.4%), or flu-like illness (6.37%) as defined by the MedDRA lower level term. Figure 2 shows representative profiles over time-in-trial and calendar-time, whereas Figure 3 shows the proportion of the participants in the clinical trial within each state over time-in-trial. The majority of the RTI state datapoints were mild ($n = 1606$, 70.9%) or moderate ($n = 611$, 27%), whereas a state of severe and higher together accounted

for 49 of the data points (2.16%; Table S1). Of the individuals with RTIs, 246 (42.9%) had more than one RTI event, with only 17 (2.96%) having overlapping RTI events which were therefore integrated within the analysis and no inter-event variability was characterized.

Model development

The structural Markov Chain model consisted of four states for healthy (HS 0), mild (HS 1), moderate (HS 2), and severe and higher lumped into a single state (HSs 3–7) given their limited data (Figure 1b). Lumping participants who died into a state from which they could theoretically transition to other states was, although counterintuitive, considered the best, parsimonious approach, in comparison to adding an additional state with only two observations to support, or removing those individuals from the analysis. The IIV model contained IIV terms on the transition probabilities $P_{2 \rightarrow 1}$, $P_{2 \rightarrow 0}$, and $P_{1 \rightarrow 0}$, with the following parameter-covariate relationships: infection pressure on $P_{0 \rightarrow 1}$, $P_{0 \rightarrow 2}$, $P_{0 \rightarrow 3}$, and $P_{2 \rightarrow 1}$; SARS-CoV-2 serology on $P_{0 \rightarrow 1}$ and $P_{0 \rightarrow 2}$; ethnicity on $P_{1 \rightarrow 0}$; biological gender on $P_{1 \rightarrow 0}$; baseline SARS-CoV-2 serology on $P_{2 \rightarrow 0}$, and SARS-CoV-2 or influenza vaccination on $P_{1 \rightarrow 0}$ (Table 1). The covariate relationships will be described in more detail below. Given that 70.9% of datapoints were for mild RTI events, whereas 27% and 2.16% were for moderate, or severe and higher RTI events, respectively, covariate relationships for transitions relating to the latter two states are to be interpreted carefully.
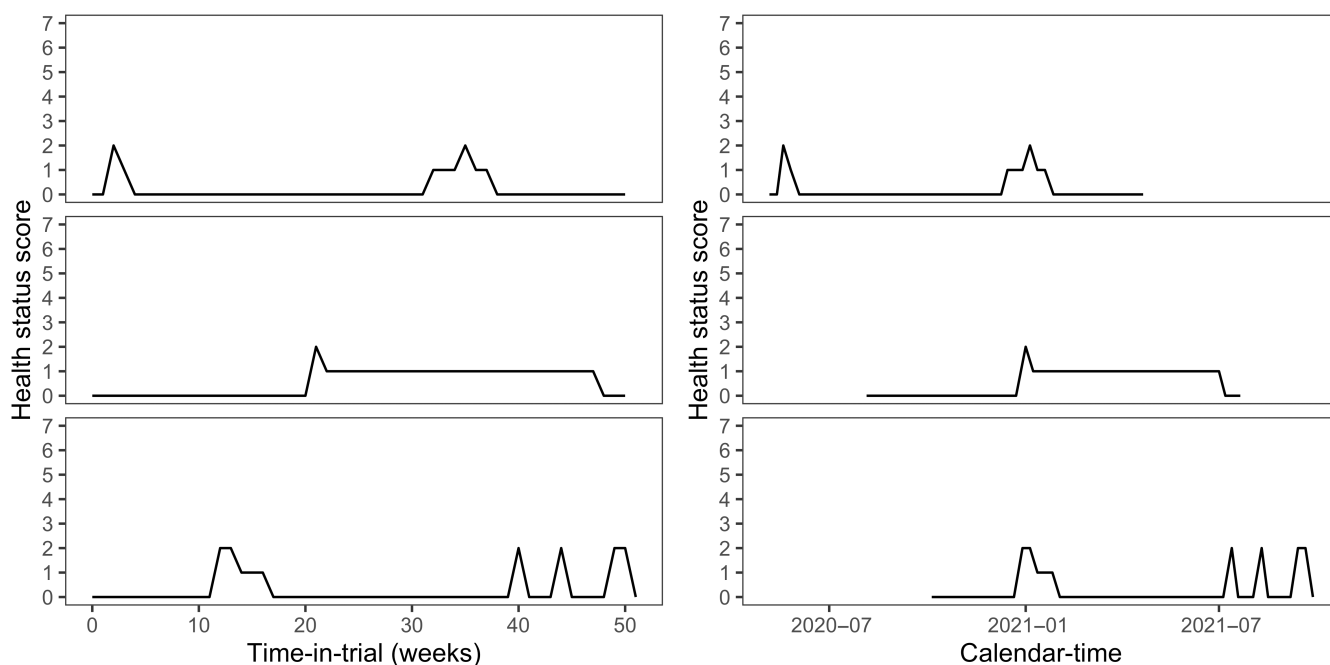

FIGURE 2 Respiratory tract infection health score over time-in-trial (left) and calendar-time (right) profiles for three individuals.
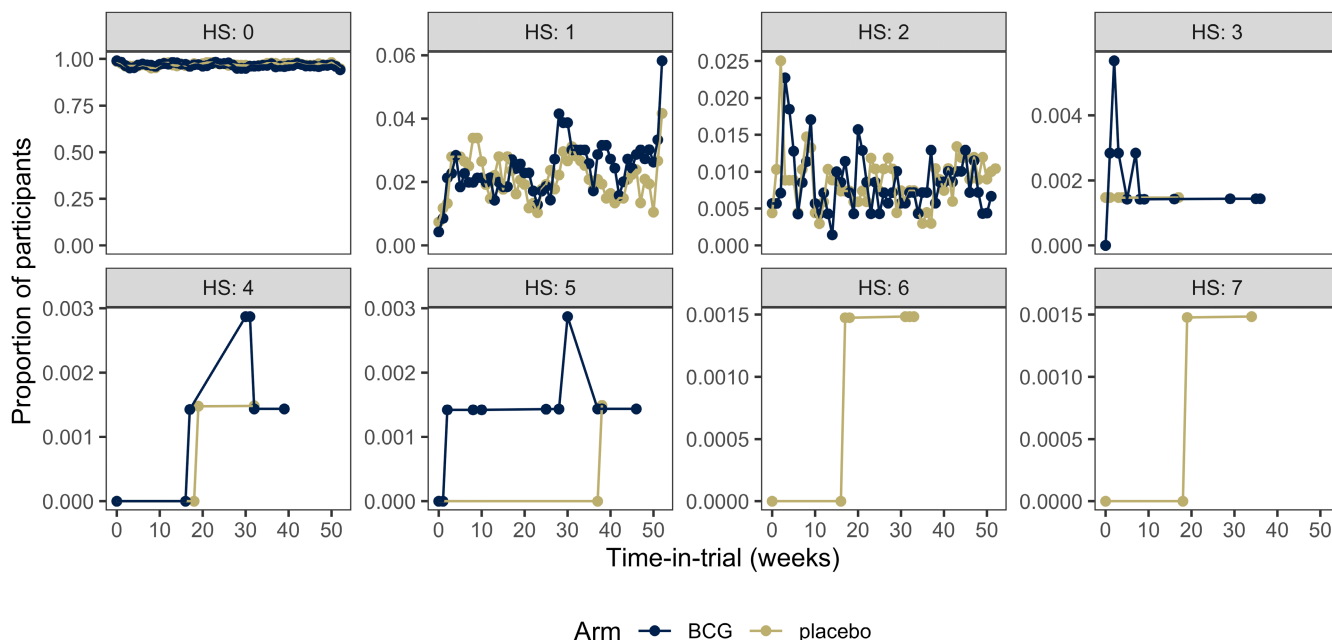
FIGURE 3 Proportion of participants in the clinical trial in each health score (HS) per study arm. BCG, Bacillus Calmette-Guérin.

Epidemiology-informed infection pressure was best described by a linear function on $P_{0\to1}$, $P_{0\to2}$, $P_{2\to1}$ (Equation 4), and by an exponential function on $P_{0\to3}$ (Equation 5).

$$\mathrm{PL}_{i\to j,k} = \mathrm{PL}_{i\to j} + \eta_k + \mathrm{CC} \cdot \mathrm{slope} \tag{4}$$

$$\mathrm{PL}_{i\to j,k} = \mathrm{PL}_{i\to j} + \eta_k + \left(\frac{\mathrm{CC}}{\mathrm{median(CC)}}\right)^{\mathrm{exponent}} \tag{5}$$

where $\mathrm{PL}_{i\to j,k}$ is the logit transformed transition probability from state i to state j for individual k conditional on not transitioning to another state, $\mathrm{PL}_{i\to j}$ is the logit transformed population transition probability from state i to state j conditional on not transitioning to another state, η_k is the individual deviation from the population parameter value for individual k drawn from a normal distribution with mean 0 and variance ω^2, and CC is the number of COVID-19 cases per capita at that calendar date.

Probabilities of transitioning to a higher HS, that is, becoming ill or worsening in illness, increased with the infection pressure of increasing COVID-19 cases (i.e., an active pandemic wave), whereas the probability of transitioning to a lower HS, that is, recovery, had a negative slope and therefore decreased with COVID-19 waves. Any positive SARS-CoV-2 antibody serology during the trial corresponded to a decreased probability of developing mild or moderate RTI symptoms. The probability of going from healthy to mild RTI symptoms decreased 23.1%, and from healthy to moderate RTI symptoms decreased 50.1%.

A higher probability of recovery ($P_{1\to0}$) was found for African ethnicity and male gender corresponding to a 16.7 and 22.7 percentage point increase in transition probability compared to non-Africans (e.g., from 37.6% to 54.3%), and women, respectively. Participants with SARS-CoV-2 seropositivity at baseline had a higher probability of recovery from a moderate RTI state ($P_{2\to0}$) than SARS-CoV-2 serology negative participants, corresponding to a 29.1 percentage point increase in transition probability. SARS-CoV-2-specific or influenza vaccination was associated with a lower probability of transitioning from mild symptoms to healthy. With increased probabilities for transitions to a lower HS, African ethnicity, male biological gender, and SARS-CoV-2 seropositivity at any time were protective against RTI symptom development and worsening, and lead to a higher probability of RTI symptom relief.

Diagnostics of the final covariate model showed, in general, acceptable model performance (Figure S1). However, the calendar-time diagnostics showed signs of misfitting the first part of the trial which coincided with the first COVID-19 wave in South Africa (Figure 4). Correspondingly, the time-in-trial diagnostics showed the predicted proportion of participants in HS 0 going down over the full-time course without stabilizing which was unexpected, with a similar upward trend for HS 1 (Figure S1, left panel), both of which were not clear in the observations. The different diagnostics suggested to have a closer look at the first period of the trial. From the COVID-19 waves in Figure 4, it was clear that the first wave of reported COVID-19 cases was substantially smaller than the second and third waves that coincided with the trial. Rather than a lower number of actual COVID-19 infections, it is likely the number of reported COVID-19 cases was lower, due to limited COVID-19 testing availability at the start of the pandemic. A parameter

TABLE 1 Parameter estimates for the final model.

Parameter	Estimate	Precision (RSE %)
Structural parameters		
$T_{0\to0}$ (%)	98.5	–
$T_{0\to1}$ (%)	0.918	7
$T_{0\to2}$ (%)	0.593	8
$T_{0\to3-7}$ (%)	0.00611	68
$T_{1\to0}$ (%)	40.2	8
$T_{1\to1}$ (%)	58.3	–
$T_{1\to2}$ (%)	1.17	24
$T_{1\to3-7}$ (%)	0.274	68
$T_{2\to0}$ (%)	28.2	10
$T_{2\to1}$ (%)	43.3	10
$T_{2\to2}$ (%)	27.1	–
$T_{2\to3-7}$ (%)	1.36	76
$T_{3-7\to0}$ (%)	4.35	103
$T_{3-7\to1}$ (%)	43.5	26
$T_{3-7\to2}$ (%)	26.1	23
$T_{3-7\to3-7}$ (%)	26.1	–
Covariates (log odds)		
Seasonal influence slope on $P_{0\to1}$	1480	21
Seasonal influence slope on $P_{0\to2}$	2580	13
Seasonal influence exponent on $P_{0\to3}$	0.472	20
Seasonal influence slope on $P_{2\to1}$	−1470	66
SARS-CoV-2 or influenza vaccination on $P_{1\to0}$	−0.592	37
SARS-CoV-2 serology on $P_{0\to1}$	−0.265	49
SARS-CoV-2 serology on $P_{0\to2}$	−0.699	25
African ethnicity on $P_{1\to0}$	0.67	36
Male biological gender on $P_{1\to0}$	0.883	26
SARS-CoV-2 serology at baseline on $P_{2\to0}$	1.23	41
COVID-19 under-reporting	2.76	23
Treatment effect (log odds)		
BCG on $P_{0\to3}$	0.804	73
BCG on $P_{2\to3}$	1.17	78
IIV [shrinkage]		
$IIV_{1\to0}$ (% coefficient of variation)	133	9 [65]
$IIV_{2\to0}$ (% coefficient of variation)	136	20 [74]
$IIV_{2\to1}$ (% coefficient of variation)	131	22 [78]

Note: $T_{x\to y}$ shows the probability from state x to y, $P_{x\to y}$ shows the probability from state x to y conditional on nontransitioning to another state. Probability to remain within the state (e.g., $T_{0\to0}$) is inferred so no relative standard error is reported. Transition probabilities originating from the same state sum up to 100%.

Abbreviations: BCG, Bacillus Calmette-Guérin; COVID-19, coronavirus disease 2019; IIV, interindividual variability; RSE, relative standard error; SARS-CoV-2, severe acute respiratory syndrome-coronavirus 2.

to estimate the fold-change in COVID-19 reporting during the first peak compared to the later peaks, was tested and included in the final covariate model. By including this element into the model, the epidemiological data on the infection pressure informed on RTIs while simultaneously, the RTI data informed on the epidemiological situation which was limited by the resources available at that time. The model estimated that COVID-19 cases were likely under-reported 2.76-fold. Inclusion of this parameter accounting for the under-reporting was statistically significant (dOFV = −30.1, p = 0.0000000407) and improved the diagnostics substantially regarding the misfits of the first wave in calendar-time and the unexpected trends in time-in-trial (Figure 5).

Including the effect of BCG in the final covariate model, showed a statistically significant 2.9-fold increase in moderate symptom worsening ($P_{2\to3}$; dOFV = −5.266, p value = 0.0218) and two-fold increase in development of severe symptoms ($P_{0\to3}$; dOFV = −5.051, p value = 0.0246) for the BCG group compared to placebo. With increased probabilities for transitions to a higher state, BCG appears to result in RTI symptom worsening. The final parameter estimates including precision is given in Table 1 and transition probabilities including effect thereon of statistically significant covariates are visualized in Figure 6.

DISCUSSION

RTIs are a threat to global health and are complex to study due to their multifactorial nature, including seasonal influence. The increased awareness and funding during the COVID-19 pandemic offered a unique opportunity to study RTIs systematically over an extended period of time to account for infection pressure on RTI occurrence and severity, for which we have developed a Markov Chain model.

Clinical trials studying indications or end points that are subject to seasonal influence benefit from longitudinal analysis related to both time-in-trial and calendar-time, to elucidate seasonal patterns that otherwise would remain unaccounted for. Exploratory graphics on both timescales, such as Figure 2, can guide thinking, as the RTI events for these three participants appear unrelated relative to time-in-trial, whereas they are overlapping in the calendar-time graph. The model-based approach we developed incorporated both time-in-trial, which is relevant to interpret the studied intervention, as well as calendar-time, which is relevant to study the fluctuating regional risk of RTIs based on the assumption that the COVID-19 waves were driving the RTI events more so than the influenza seasonality.[23] Indeed,
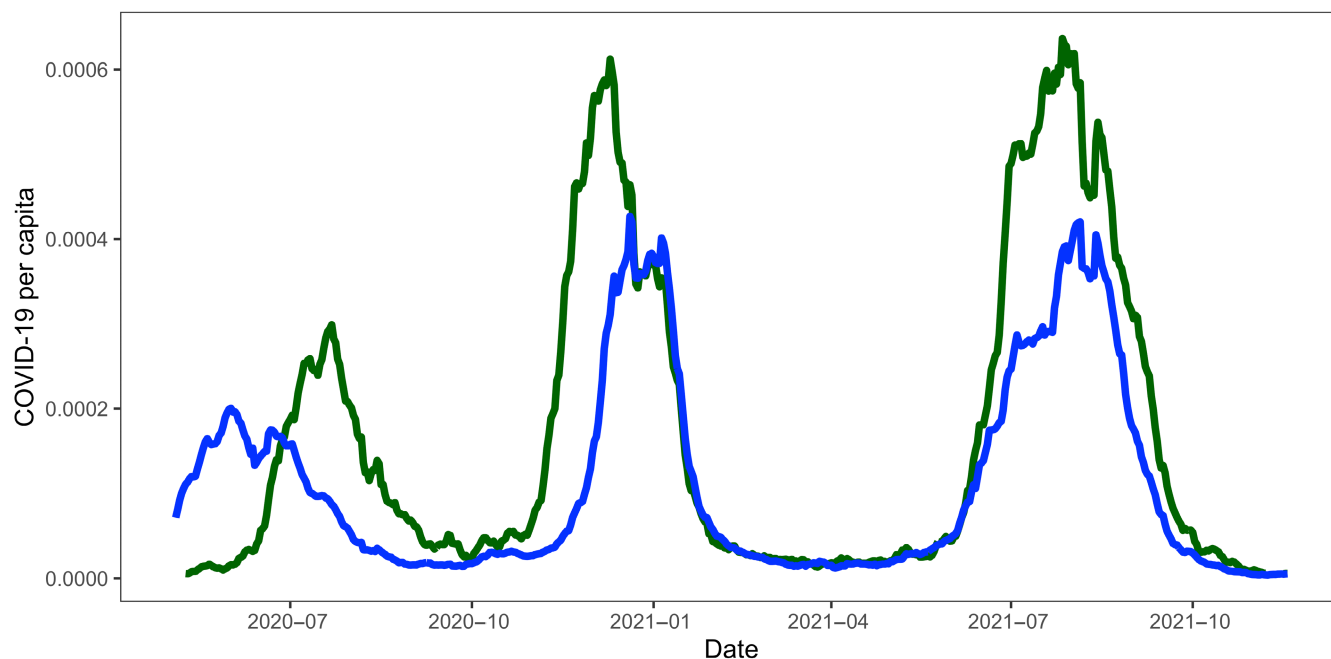
FIGURE 4 COVID-19 cases per capita in Cape Town (blue, study sites TASK Central and University of Cape Town Lung Institute) and the Garden Route (green, study site TASK Eden). COVID-19, coronavirus disease 2019.

the model found statistically significant effects on the probability of RTI symptom worsening with increasing COVID-19 cases per capita, as well as on the probability of RTI symptom relief with decreasing COVID-19 cases per capita. In addition, the relationship between the COVID-19 waves and RTI incidence, based on the full study dataset, enabled the model to estimate the fold-change in under-reporting of COVID-19 cases during the first wave in South Africa, compared to later peaks.[24] Despite the possibility it provided to estimate this under-reporting, the inclusion of epidemiology on only COVID-19 cases was a limitation of this work. The impact of infection pressure could be quantified in more detail with additional viral seasonal or non-seasonal data, or epidemiology on overall RTI cases.

The Markovian model developed here was able to elucidate covariate relationships, including risk factors for disease severity that are of clinical relevance for managing RTIs in healthcare workers, defining those healthcare workers that are at increased risk of developing RTI, and predicting subsequent consequences on their health and performance. Quantitative model-based evidence is provided for protective effect of positive serology for SARS-CoV-2 antibodies both at baseline and during the trial,

in line with expectations of infection-based immunity. Almost a quarter of the RTIs reported were COVID-19 diagnoses, likely more considering the under-reporting of COVID-19 at the start of the trial, which was quantified as well using our model-based approach. Evidence based prioritization of healthcare workers at high risk of severity, based on those without protective risk factors (biological gender, ethnicity, and serology where relevant), may support active monitoring for disease progression allowing early interventions.

A main limitation to the dataset studied was the fact that data was acquired through symptom reporting by the participants, which may introduce bias. No SARS-CoV-2 or other pathogen virology testing was mandated in the study, and infections that were asymptomatic or that participants did not report, might have been missed. For example, within participants that seroconverted from no SARS-CoV-2 antibodies to detectable SARS-CoV-2 antibodies, 35% did not report a COVID-19 or RTI event in general during the conversion interval. In the 65% that did report an event, the majority were COVID-19 (74%), as expected. Serology data was sparsely only available and no clear correlation was detected between seroconversion and SARS-CoV-2 specific vaccination. Despite this

FIGURE 5 Simulation based diagnostics of the final covariate model of proportion of participants of the trial in health score (HS) 0 (healthy), 1 (mild symptoms), 2 (moderate symptoms), and 3–7 (severe symptoms and up) over time-in-trial (a) and calendar-time (b), and of the proportion of participants with the corresponding transition (c). Shaded area shows the 95% simulation interval, solid line shows the median from the simulation, and symbols are the observed proportions.
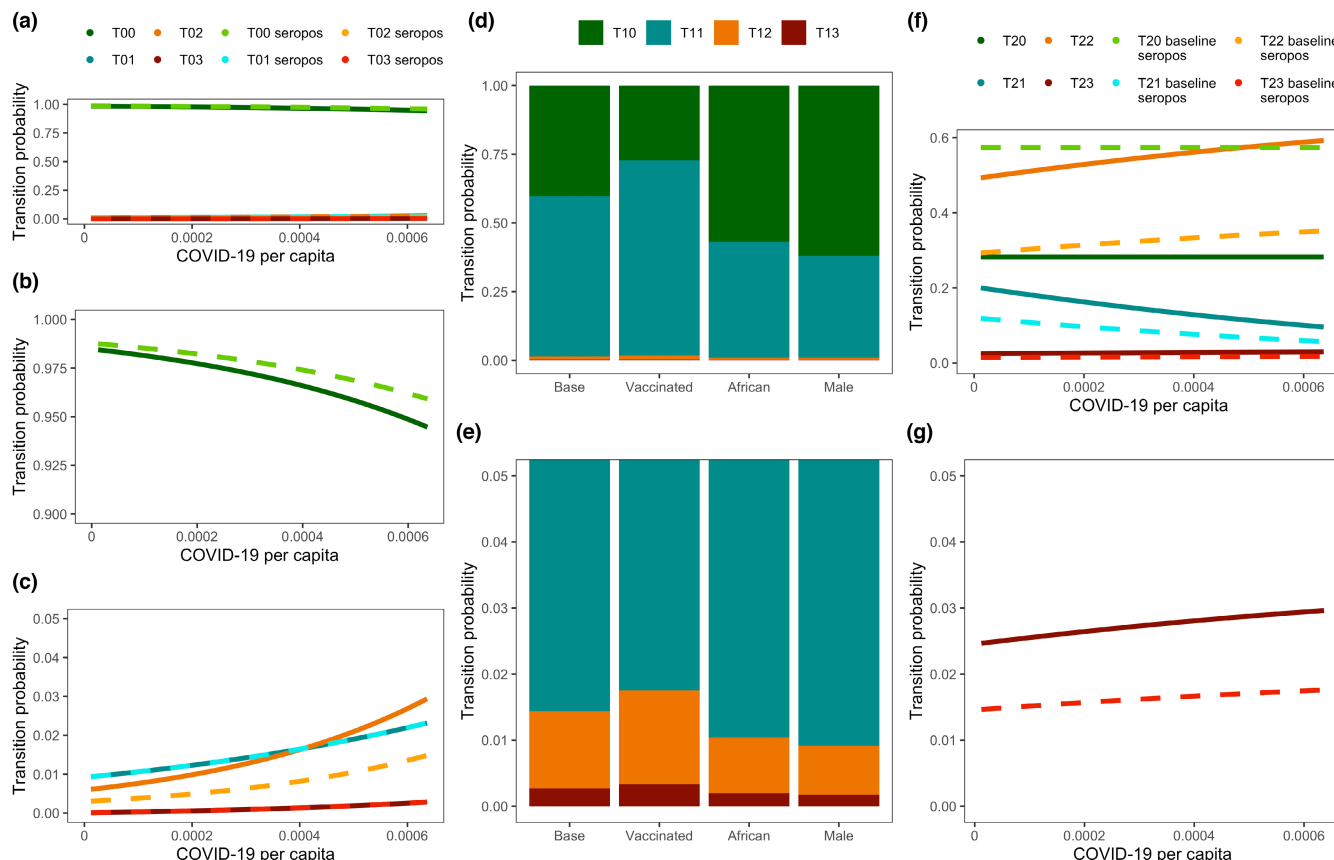
(a)

(b)

(c)
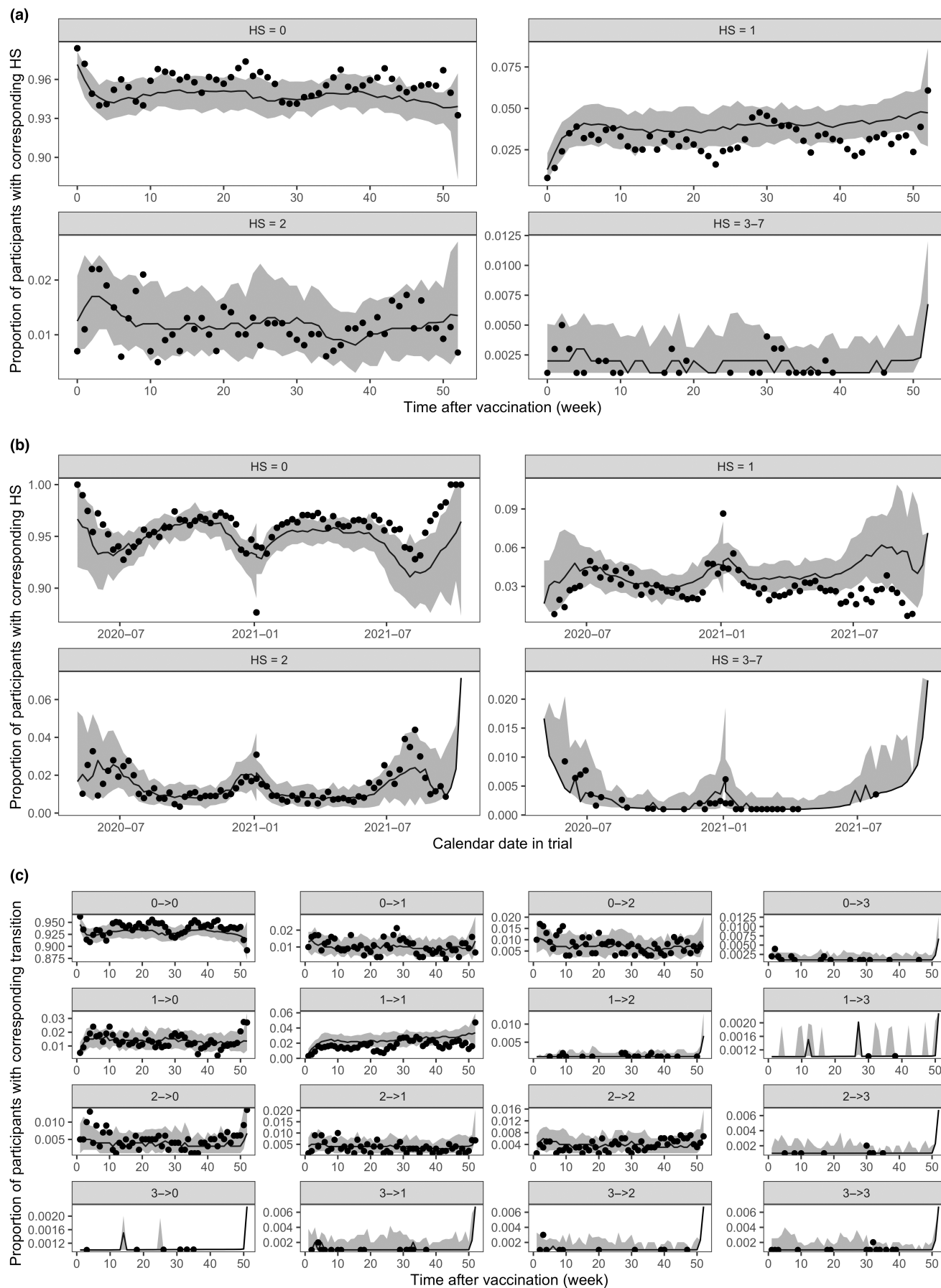
FIGURE 6 Transition probabilities and the impact of covariate relationships. T_{xy} shows the transition probability of state x to state y. (a–c) Transition probabilities from healthy state (HS) 0 as a function of infection pressure of COVID-19 cases alone (solid line) and combined with SARS-CoV-2 seropositivity at any timepoint (dashed line), with zoomed panel for clearer visualization for transition probabilities 0.9–1 (b) and 0–0.05 (c). (d–e) Transition probabilities from mild symptoms state 1 over identified covariate relationships of vaccination, ethnicity, and gender, compared to their control (base), with zoomed panel for clearer visualization for transition probabilities 0–0.05 (e). (f–g) Transition probabilities from moderate state 2 as a function of infection pressure of COVID-19 cases alone (solid line) and combined with SARS-CoV-2 seropositivity at baseline (dashed line), with zoomed panel for clearer visualization for transition probabilities 0–0.05 (g). COVID-19, coronavirus disease 2019; SARS-CoV-2, severe acute respiratory syndrome-coronavirus 2.

reporting bias, the presented analysis still has value on RTIs that were serious enough to report. The bias could be a reason why men had a higher probability of recovery from RTI symptoms, whereas they are generally at higher risk.[25] Unexpectedly, African ethnicity were also found to have a higher probability of recovery from RTI symptoms, although they have been reported to be risk factors for severe COVID-19.[26] The trial population of healthcare workers here perhaps is a confounding factor, as the suggested mechanism of inequal access to healthcare for the full population would apply less to healthcare workers studies here.[26] It was surprising that SARS-CoV-2-specific and influenza vaccination decreased the probability of transitioning from mild symptoms to healthy. This might be an artifact in the model, as only a small minority of these transitioning events (13%) were in vaccinated people. A sensitivity analysis of constraining the model to a protective effect of the SARS-CoV-2 or influenza vaccination did not result in the selection of this

covariate. BCG (re)vaccination did not protect against COVID-19 or RTIs in general, which concurs with the primary analysis of this trial.[11] Another limitation of the study was the limited amount of data for moderate, or severe and higher RTI events, respectively, consisting of only 27% and 2.16% of datapoints, respectively, with its consequence for interpreting the corresponding covariate relationships for transitions relating to those two states. This includes the quantified intervention relationship of BCG vaccination, which could explain its large imprecision and therefore should be interpreted with care. Lastly, this clinical trial was performed in the initial phase of the COVID-19 pandemic in South Africa with all corresponding limitations to diagnostic and preventive possibilities, and presented results should be interpreted within that context. In conclusion, a Markov Chain model was developed to quantify the onset and probability of symptom worsening and relief as a function of time, where risk factors included infection pressure, and BCG (re)

vaccination. The modeling workflow developed here included time-in-trial and calendar-time elements in model development and diagnostics, which should be applied when studying indications, end points, and/or interventions which are subject to time varying fluctuations, such as seasonal influence.

AUTHOR CONTRIBUTIONS

R.C.vW., L.M., C.M.U., J.P., A.H.D., and U.S.H.S. wrote the manuscript. R.C.vW., L.M., C.M.U., J.P., A.H.D., and U.S.H.S. designed the research. R.C.vW., L.M., C.M.U., J.P., A.H.D., and U.S.H.S. performed the research. R.C.vW. and U.S.H.S. analyzed the data.

FUNDING INFORMATION

CONFLICT OF INTEREST STATEMENT

The authors declared no competing interests for this work.

ORCID

Rob C. van Wijk https://orcid.org/0000-0001-7247-1360
Laurynas Mockeliunas https://orcid.org/0000-0002-7008-4946
Ulrika S. H. Simonsson https://orcid.org/0000-0002-3424-9686

REFERENCES

1. Centers for Disease Control and Prevention Disease burden of influenza. 2022. https://www.cdc.gov/flu/about/disease/burden.htm.
2. Ferkol T, Schraufnagel D. The global burden of respiratory disease. *Ann Am Thorac Soc.* 2014;11:404-406. doi:10.1513/AnnalsATS.201311-405PS
3. Miroyama M, Hugentobler WJ, Iwasaki A. Seasonality of respiratory viral infections: will COVID-19 follow suit? *Ann Rev Virol.* 2020;7:87-101. doi:10.3389/fpubh.2020.567184
4. Feldman C, Shaddock E. Epidemiology of lower respiratory tract infections in adults. *Expert Rev Respir Med.* 2019;13:63-77. doi:10.1080/17476348.2019.1555040
5. Obando-Pacheco P, Justicia-Grande AJ, Rivero-Calle I, et al. Respiratory syncytial virus seasonality: a global overview. *J Infect Dis.* 2018;217:1356-1364. doi:10.1093/infdis/jiy056
6. Cherednik I. Modeling the waves of Covid-19. *Acta Biotheor.* 2022;70:8. doi:10.1007/s10441-021-09428-w
7. Chowell G, Tariq A, Hyman JM. A novel sub-epidemic modeling framework for short-term forecasting epidemic waves. *BMC Med.* 2019;17:1-18. doi:10.1186/s12916-019-1406-6
8. He D, Chiu APY, Lin Q, Cowling BJ. Differences in the seasonality of Middle East respiratory syndrome coronavirus and influenza in the Middle East. *Int J Infect Dis.* 2015;40:15-16. doi:10.1016/j.ijid.2015.09.012
9. Ayala A, Villalobos Dintrans P, Elorrieta F, Castillo C, Vargas C, Maddaleno M. Identification of COVID-19 waves: considerations for research and policy. *Int J Environ Res Public Health.* 2021;18:11058. doi:10.3390/ijerph182111058
10. Kamal MA, Smith PF, Chaiyakunapruk N, et al. Interdisciplinary pharmacometrics linking oseltamivir pharmacology, influenza epidemiology and health economics to inform antiviral use in pandemics. *Br J Clin Pharmacol.* 2017;83:1580-1594. doi:10.1111/bcp.13229
11. Upton CM, van Wijk RC, Mockeliunas L, et al. Safety and efficacy of BCG Re-vaccination in reducing COVID-19 morbidity in healthcare workers: a double-blind, randomised, controlled, phase 3 trial. *EClinicalMedicine.* 2022;48:101414. doi:10.1016/j.eclinm.2022.101414
12. Olsen SJ, Azziz-Baumgartner E, Budd AP, et al. Decreased influenza activity during the COVID-19 pandemic — United States, Australia, Chile, and South Africa, 2020. *Morb Mortal Wkly Rep.* 2020;69:1305-1309. doi:10.1111/ajt.16381
13. Western cape government health department circular H 102/2020 Covid-19 population data. Chief Director: Strategy and Health Support 1–123. 2020 https://www.westerncape.gov.za/assets/departments/health/h_102_2020_covid-19_population_data.pdf
14. Lacroix BD, Lovern MR, Stockis A, Sargentini-Maier ML, Karlsson MO, Friberg LE. A pharmacodynamic markov mixed-effects model for determining the effect of exposure to certolizumab pegol on the ACR20 score in patients with rheumatoid arthritis. *Clin Pharmacol Ther.* 2009;86:387-395. doi:10.1038/clpt.2009.136
15. Jonsson EN, Karlsson MO. Automated covariate model building with NONMEM. *Pharm Res.* 1998;15:1463-1468. doi:10.1023/a:1011970125687
16. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr.* 1974;19:716-723. doi:10.1109/TAC.1974.1100705
17. Nguyen THT, Mouksassi MS, Holford N, et al. Model evaluation of continuous data pharmacometric models: metrics and graphics. *CPT Pharmacometrics Syst Pharmacol.* 2017;6:87-109. doi:10.1002/psp4.12161
18. Beal S, Sheiner L, Boeckmann A, Bauer RJ, eds. *NONMEM 7.5.0 Users Guides.* ICON Development Solutions; 1989-2020.
19. Lindbom L, Pihlgren P, Jonsson N. PsN-toolkit – a collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. *Comput Methods Programs Biomed.* 2005;79:241-257. doi:10.1016/j.cmpb.2005.04.005
20. R statistical computing and graphics software environment. https://www.r-project.org/
21. Van Wijk RC, Mockeliunas L, Simonsson USH. Reproducibility in pharmacometrics. Github Repository Website. 2023. https://github.com/rcvanwijk/ReproducibilityPMX
22. RStudio Team RStudio: Integrated Development for R. 2016. http://www.rstudio.com/
23. Tempia S, Walaza S, Bhiman JN, et al. Decline of influenza and respiratory syncytial virus detection in

facility-based surveillance during the COVID-19 pandemic, South Africa, January to October 2020. *Euro Surveill*. 2021;26:1-10. doi:10.2807/1560-7917.ES.2021.26.29.2001600

24. Gu X, Mukherjee B, Das S, Datta J. Covid-19 prediction in South Africa: estimating the unascertained cases- the hidden part of the epidemiological iceberg. *medRxiv* 2020.12.10.20247361. 2021. doi:10.1101/2020.12.10.20247361

25. Falagas ME, Mourtzoukou EG, Vardakas KZ. Sex differences in the incidence and severity of respiratory tract infections. *Respir Med*. 2007;101:1845-1863. doi:10.1016/j.rmed.2007.04.011

26. Jassat W, Cohen C, Tempia S, et al. Risk factors for COVID-19-related in-hospital mortality in a high HIV and tuberculosis prevalence setting in South Africa: a cohort study. *Lancet HIV*. 2021;8:e554-e567. doi:10.1016/S2352-3018(21)00151-X

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.